



TITLE:

Method for Energy Computation of Protein Structures : A Set of Amino Acid Geometries and Energy Parameters (Commemoration Issue Dedicated to Professor Tatsuo Yamamoto on the Occasion of his Retirement)

AUTHOR(S):

Nishikawa, Ken; Oobatake, Motohisa; Ooi, Tatsuo

CITATION:

Nishikawa, Ken ...[et al]. Method for Energy Computation of Protein Structures : A Set of Amino Acid Geometries and Energy Parameters (Commemoration Issue Dedicated to Professor Tatsuo Yamamoto on the Occasion of his Retirement). Bulletin of the Institute for Chemical Research, Kyoto University 1980, 58(3): 308-322

ISSUE DATE:

1980-05-31

URL:

<http://hdl.handle.net/2433/76901>

RIGHT:

Method for Energy Computation of Protein Structures. A Set of Amino Acid Geometries and Energy Parameters

Ken NISHIKAWA, Motohisa OOBATAKE, and Tatsuo OOI*

Received March 19, 1980

A complete set of amino acid geometries and the energy functions for the computation of the conformational energy on the basis of an X-ray structure of a protein is presented. Our method is based on the rigid-geometry treatment, i.e., all the bond lengths and bond angles are fixed, and the planer-peptide geometry is assumed, thus leaving internal rotations about the single bonds as independent variables. We consider nonbonded, torsional and electrostatic interactions and a loop-closing potential of the disulfide bond for contributions to the total conformational energy of a protein. No solvent effects are included explicitly. Our method may have the following unique aspects; the expression of the hydrogen bond energy as a combination of the electrostatic interaction and the nonbonded energy, allowance of the flexibility and associated ring-deformation potential for proline ring, and also the separate generation of the backbone conformation and subsequent attachment of side chains. A method developed to calculate the first derivative of conformational energies is also presented.

KEY WORDS: X-ray coordinates/ energy calculation/ refinement/

INTRODUCTION

Three dimensional structures of more than fifty different kinds of proteins have been known in the past decade by X-ray crystallography. The atomic coordinate data of proteins stored in Protein Data Bank¹⁾ have been used for a wide variety of analyses of protein structures.²⁾ The experimental coordinates (usually around 2Å resolution) are, however, not so accurate for some purposes, particularly to examine the protein stability in terms of energy. Since the interaction energy involved in a protein is highly sensitive to the position of individual atoms, it is rather usual to find several severe atomic overlaps with use of the raw data of X-ray coordinates. Hence, the refinement of protein coordinates is necessary prior to the energetic analysis of a protein structure performed.

Several different procedures for the energetic refinement have been developed so far.³⁻⁵⁾ The characteristic features of the method which has been developed in this laboratory⁵⁾ are to use the rigid geometry (the standard bond lengths and bond angles adopted) of the peptide unit and side chains in contrast to the method with flexible geometry,³⁾ and to fit the overall protein conformation to the X-ray structure by using the relative distances between residue pairs as a guide in contrast to the usage of the absolute atomic coordinates as a guide.⁴⁾ The latter point implies that our fitting procedure does not depend on the coordinate systems to represent both the X-ray and calculated structures, so that the computational method could be

* 西川 建, 大島玄久, 大井龍夫: Laboratory of Physical Chemistry of Enzyme, Institute for Chemical Research, Kyoto University, Uji, Kyoto.

simplified.

Our method of energetic refinement and its application to bovine pancreatic trypsin inhibitor (BPTI) were described elsewhere.^{5b)} It has been found in that study^{5b)} that the low energy structure is restricted as expected but has considerable range of freedom in a sense. That is, the concerted motion of variables (the backbone dihedral angles) with keeping the low level of the total energy and also fitting the overall conformation to the X-ray structure is possible. This freedom must correspond to the flexibility which the real native protein possesses in solution.⁶⁾

In this paper, we will present the whole set of geometric and energetic parameters for conformational energy calculation, and some new devices for the chain generation and energy minimization processes.

GEOMETRY

In our computation, all the bond lengths and bond angles of a polypeptide chain are fixed at their standard values, so that variables to describe the conformation are dihedral angles around the single covalent bonds. The planer peptide group is also assumed to fix the rotation about the C'-N bond (ω) in the *trans* configuration except for proline, for which either *trans* or *cis* configuration is chosen in accordance with that of the X-ray structure. Since hydrogen atoms are not recognizable usually by the X-ray diffraction technique, it is reasonable to incorporate them into so-called united atoms⁷⁾: for instance, the methylene group is treated as one atom which has a greater van der Waals radius than a single carbon atom. However, those hydrogen atoms which could participate in hydrogen bonding, e.g., H atoms of amino and carboxyl groups, are treated explicitly in this study because the position of a hydrogen atom is crucial for the directionality of a hydrogen bond. As a result, all the atoms considered for the backbone peptide unit are C $^{\alpha}$, N, H, C' and O, where C $^{\alpha}$ is regarded as a united atom. The standard bond lengths and bond angles given by Pauling and Corey⁸⁾ are employed for the peptide unit as shown in Fig. 1a.

Since some freedom and variation depending on the different type of side chains are expected for the bond angle $\tau(\text{NC}^{\alpha}\text{C}')$ ⁹⁾, the use of the uniform value for this

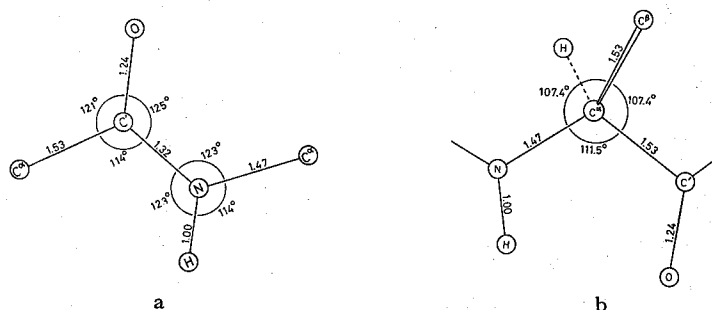


Fig. 1. Geometry of the trans-planer peptide unit (a) and the residue unit (b) of the protein backbone. The angle of $\tau(\text{NC}^{\alpha}\text{C}')$ is optional but shown here the representative value, 111.5° (see the text).

angle might be questionable. For the sake of simplicity, we use of fixed value of 111.5° as a representative one (Fig. 1b), which is close to the average value deduced from X-ray structures of several proteins.^{3a)} Other values between 109.5° (the tetrahedral angle) and 112.5° have also been attempted.^{5b)} The sensitivity of the angle to the conformation is not so significant.^{5b)}

Change of the optional value of $\tau(\text{NC}^\alpha\text{C}')$ may influence the angles of $\tau(\text{NC}^\alpha\text{C}^\beta)$ and $\tau(\text{C}'\text{C}^\alpha\text{C}^\beta)$ which determine the direction of the side-chain attachment (see Fig. 1b). The following assumptions were adopted for these bond angles. The C^β and H atoms attached to C^α (the H atom is here considered tentatively) are located on the bisecting plane of the angle $\tau(\text{NC}^\alpha\text{C}')$, and similarly N and C' are

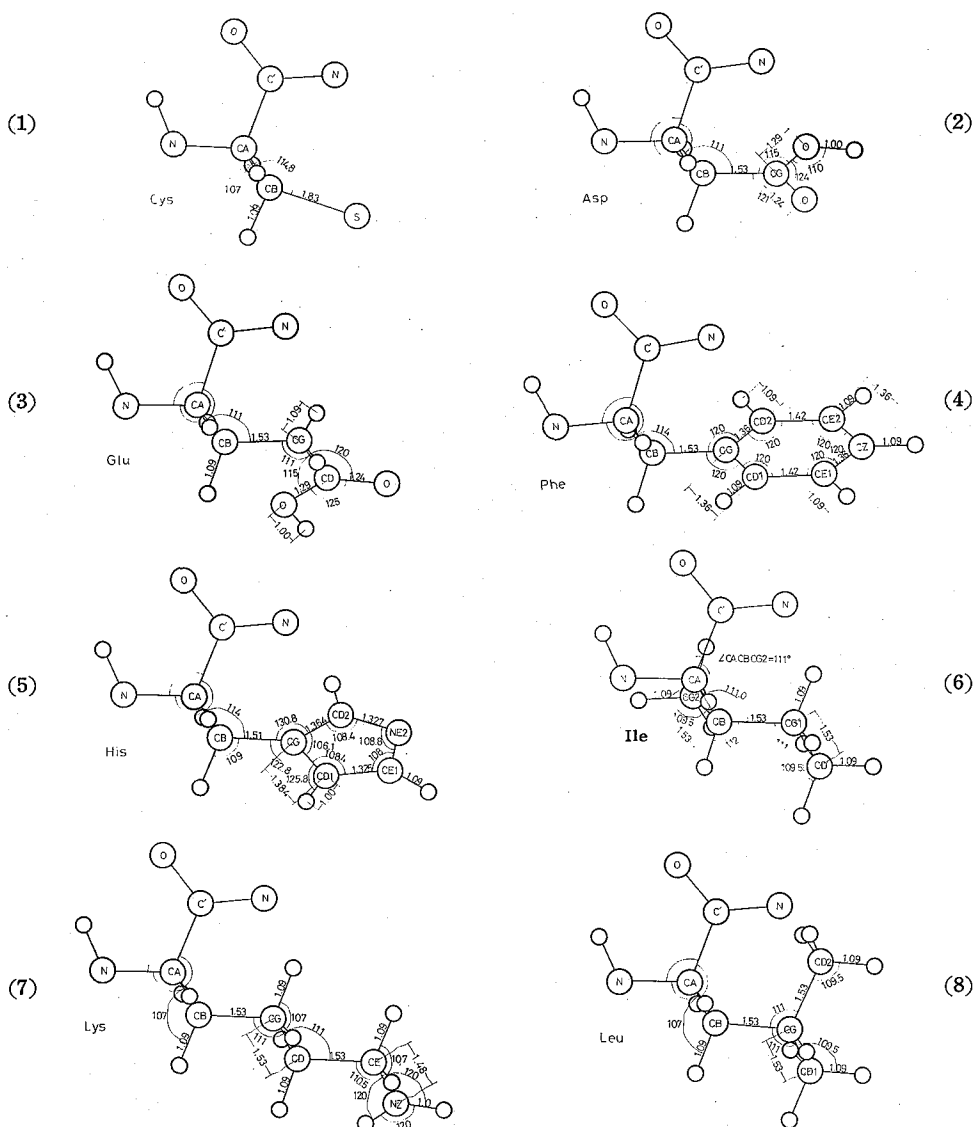


Fig. 2.

Energy Computation of Protein Structure

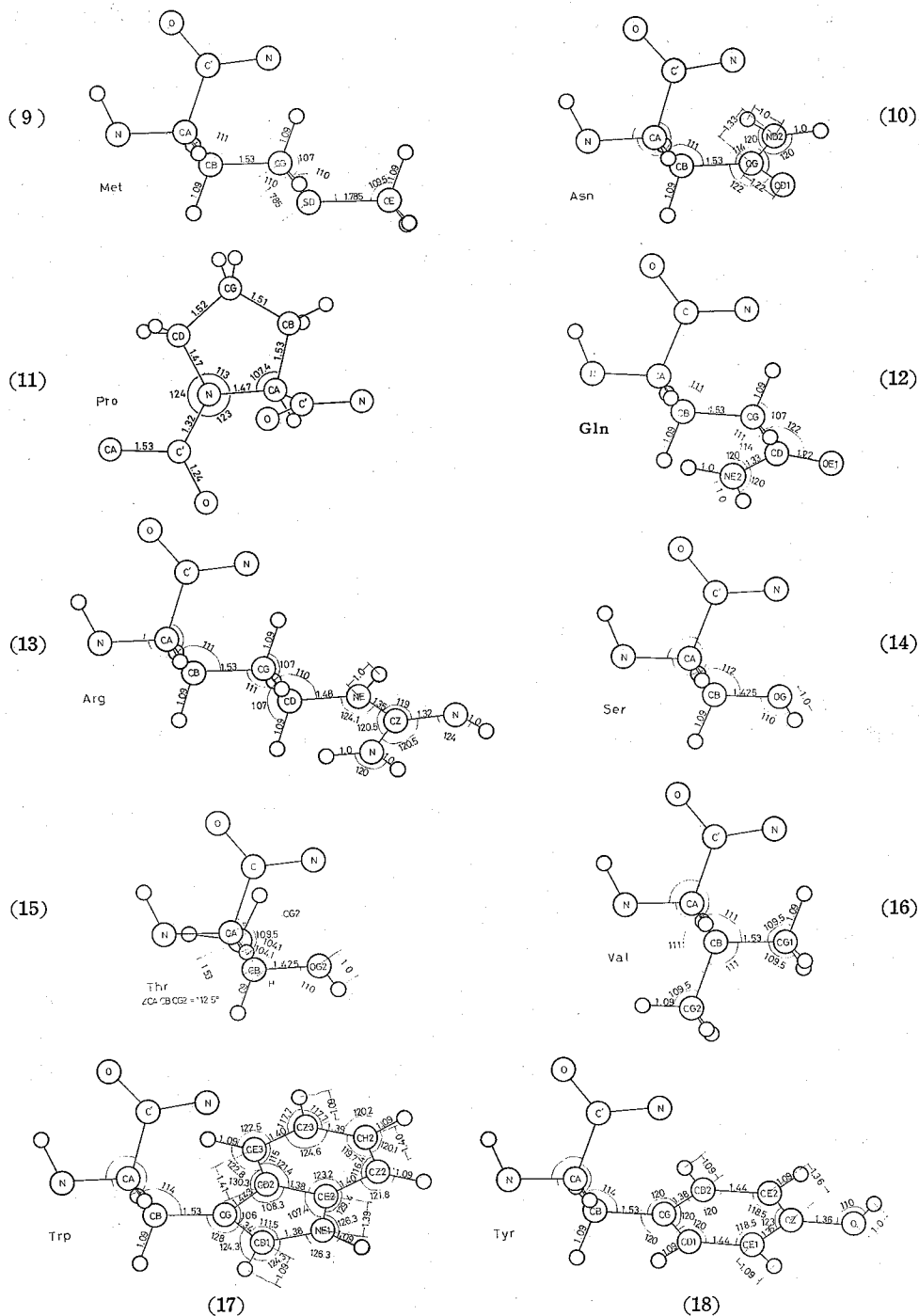


Fig. 2. Geometry of eighteen L-type amino acid residues (see Fig. 1 for Gly and Ala). All the bond angles and bond lengths of the side chains are fixed except proline, where the C7 position is not fixed due to the flexible-ring assumption (see the text). The H atoms are depicted with small circles, but nonpolar H atoms are not considered in the energy calculation since they are incorporated in united atoms.

on the bisecting plane of $\tau(\text{C}^\beta\text{C}^\alpha\text{H})$. Moreover, the angle of $\tau(\text{C}^\beta\text{C}^\alpha\text{H})$ is assumed equal to $\tau(\text{NC}^\alpha\text{C}')$. Under these conditions, both angles of $\tau(\text{NC}^\alpha\text{C}^\beta)$ and $\tau(\text{C}'\text{C}^\alpha\text{C}^\beta)$ are to be identical to each other, and its value computed against 109.5° and 111.5° of $\tau(\text{NC}^\alpha\text{C}')$ is 109.4° and 107.4° , respectively.

The geometries of all of the side chains are shown in Fig. 2. The bond lengths and bond angles are those of Momany *et al.*⁹⁾ Only an exception is for proline. While Momany *et al.*⁹⁾ treat the pyrrolidine ring as rigid and assume either of two puckered configurations of "up" and "down", we treat the ring as flexible and allow continuous deformation.^{10,11)} We assume all the bond lengths in pyrrolidine ring fixed (Fig. 2), and that the ring deformation takes place according to the dihedral angle about the N-C $^\alpha$ bond (φ)¹⁰⁾: The C $^\delta$ position is always fixed on the peptide plane of proline with the fixed bond angles of $\tau(\text{C}^\delta\text{NC}^\alpha)$ and $\tau(\text{C}^\delta\text{NC}')$, and the C $^\beta$ position in reference to the peptide plane is determined from a given value of φ . Then, the C $^\gamma$ position can be located to satisfy the fixed bond lengths of C $^\beta$ -C $^\gamma$ and C $^\gamma$ -C $^{\delta 10a}$). The ring deformation is associated with the deformation energy which is expressed as a function of the dihedral angle φ (see Section IV). Large deformation should be excluded due to corresponding high energies, so that the value of φ is consequently restricted in the limited range.

CHAIN GENERATION

With use of the fixed geometry given above, any backbone conformation of a protein is generated from a given set of (φ_i, ψ_i) , the dihedral angles about the N-C $^\alpha$ and C $^\alpha$ -C' bonds of amino acid residues, respectively. According to the general formulation,^{12a)} the generation of a backbone conformation is attained by repetitive applications of a transformation matrix $T_{i \rightarrow i-1}$ and a translation vector $p_{i \rightarrow i-1}$ for the coordinates of constituent atoms. The initial coordinates of our backbone atoms at the fully extended conformation ($\varphi = \psi = 180^\circ$)¹³⁾ are identical to those of Table II of Ref. 12a, due to the same geometry employed.

The transformation matrix for *trans* or *cis* peptide group may be expressed as follows

$$T_{i \rightarrow i-1}^{trans} = T_\alpha^Z T_{\phi+\pi}^X T_\beta^Z T_\psi^X \quad (1)$$

and

$$T_{i \rightarrow i-1}^{cis} = T_{\alpha'}^Z T_\phi^X T_\beta^Z T_\psi^X \quad (2)$$

Here, matrices T_γ^X and T_γ^Z denote the rotation by an (arbitrary) angle γ about X- and Z-axis of the coordinate system, respectively. They are defined as

$$T_\gamma^X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \gamma & -\sin \gamma \\ 0 & \sin \gamma & \cos \gamma \end{bmatrix} \quad (3)$$

and

$$T_\gamma^Z = \begin{bmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The angles α , α' and β appeared in eqs. (1) and (2) indicate the constant angles, i.e.,

$$\begin{aligned}\alpha &= \tau(\text{C}'\text{NC}^\alpha) - \tau(\text{C}^\alpha\text{C}'\text{N}) \\ \alpha' &= 2\pi - \tau(\text{C}'\text{NC}^\alpha) - \tau(\text{C}^\alpha\text{C}'\text{N}) \\ \beta &= \pi - \tau(\text{NC}^\alpha\text{C}')$$
(5)

and they are equal to 9° , 123° and 68.5° , respectively, in our geometry (with $\tau(\text{NC}^\alpha\text{C}') = 111.5^\circ$). The translation vectors $\mathbf{p}_{i \rightarrow i-1}$ defined in Ref. 12a for *trans* and *cis* peptide groups are

$$\mathbf{P}_{i \rightarrow i-1}^{\text{trans}} = \begin{bmatrix} 3.519 \\ -1.436 \\ 0.0 \end{bmatrix} \quad \text{and} \quad \mathbf{P}_{i \rightarrow i-1}^{\text{cis}} = \begin{bmatrix} 1.266 \\ -2.439 \\ 0.0 \end{bmatrix} \quad (6)$$

It should be noted that we adopt here the new convention¹³⁾ for the φ , ψ angles, which gives rise to some differences in matrix elements of eq. (1) compared with those given in Ref. 12a.

In our computation scheme for the generation of a protein, the whole conformation of the backbone is generated first with the procedure described above, and the side chains whose conformations are specified with dihedral angles of χ 's are attached to the backbone later on in a separate step. The step to generate a backbone conformation should include the C^β atom as well as the other backbone atoms for residues other than glycine, since the C^β position depends only on the φ , ψ angles. The initial coordinates of the C^β atom are calculatable depending on a given angle of $\tau(\text{NC}^\alpha\text{C}')$ as already mentioned (Fig. 1b).

A side chain of any particular type (other than alanine whose methyl group has been given by the united atom at the C^β position, and proline, the pyrrolidine ring of which is treated separately as described before) is generated by the general transformation method^{12a)} applied to the initial coordinates of all the constituent atoms in the fully extended conformation ($\chi_j = 180^\circ$). All the side-chain coordinates of i -th residue, after transformed with use of a given set of $\{\chi_j\}$, can be expressed in a *local* coordinate system defined on a plane involving the N, C^α , and C^β atoms of the i -th residue^{12a)}: the origin of the coordinate system is at the C^α atom, its x-axis is along the C^α - C^β bond and the N atoms has a positive y-coordinate. These three atoms, on the other hand, have already been expressed in a certain *global* coordinate system for the backbone. Then, it is generally possible to find such a transformation matrix that brings one coordinate system onto the other in a direct manner, if the coordinates of any three points are known in the both systems. Application of this direct transformation method enables us to attach a side chain expressed in a *local* system onto the backbone given in a *global* coordinate system.

There may be some advantage in our procedure treating a backbone and side chains separately; e.g., this makes it easier to change the side-chain conformation in keeping a fixed backbone conformation. The direct transformation method is also used in our energy minimization to compute the first derivative of the energy with respect to φ and ψ .

ENERGIES

Contributions considered to the total conformational energy are the nonbonded, torsional and electrostatic energies. Hydrogen-bond energy is treated as a special case of the electrostatic interaction as described below. Furthermore, a special potential for the deformation of proline ring and a loop-closing (artificial) potential of disulfide bond(s) if exists should also be added. However, we ignore the solvent (hydrophobic) effects. This point will be discussed in Section VI.

Our main concern in the energetic refinement of a protein structure is to find such protein conformations that possess sufficiently low energies, i.e., devoid of high energy loss arising from severe atomic overlaps, and also that are close enough to the X-ray structure of the protein as well. In this respect, the most important energy is the nonbonded (van der Waals) interaction between atom pairs.

The nonbonded energy is calculated between all the atom pairs except those of fixed distances such as the nearest and the next-nearest neighbor atoms along the covalent bonding. The potential is assumed to have the form of the Lennard-Jones (6-12) type:

$$U_{nb} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} = e_{m,ij} \left\{ \left(\frac{r_{m,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{m,ij}}{r_{ij}} \right)^6 \right\} \quad (7)$$

where r_{ij} is the distance between the atoms i and j , and the parameters of $r_{m,ij}$ and $e_{m,ij}$ denote the distance and the energy at the potential minimum, respectively. The parameter sets of (A_{ii}, B_{ii}) and $(r_{m,ii}, e_{m,ii})$ for the atompairs of identical types are listed in Table I. This parameter set is the same as that used previously.¹⁴⁾ Note that H atoms are only considered, in our calculation, for those that can participate the hydrogen bond and the other non-polarized H atoms are included in the united atoms listed. Parameter values between the pairs of hetero-atoms are derived¹⁴⁾ from the combination of the parameters given in Table I.

Table I. Nonbonded energy parameters of eq. (7), only for atompairs of identical types

| Atom | r_m (Å) | e_m (Kcal/mol) | A | B |
|---|-----------|------------------|----------|------|
| C (Carbonyl and guanidyl) | 3.40 | -0.120 | 286,000 | 370 |
| CH, CH ₂ , CH ₃ (Aliphatic) | 4.00 | -0.128 | 2140,000 | 1040 |
| C, CH (Aromatic) | 3.80 | -0.144 | 1300,000 | 864 |
| O | 3.04 | -0.232 | 145,000 | 367 |
| N | 3.10 | -0.205 | 161,000 | 363 |
| NH (Aromatic) | 3.40 | -0.218 | 520,000 | 673 |
| S | 3.60 | -0.091 | 433,000 | 398 |
| H | 2.40 | -0.123 | 4,490 | 47 |

The intrinsic torsional potential for the backbone rotational bonds is suggested to have only low barrier heights and sometimes has been neglected.⁹⁾ We employ the following potentials adopted by Scheraga^{12c)} (replaced to the new convention¹³⁾ of φ and ψ)

Energy Computation of Protein Structure

$$U(\varphi) = (U_\varphi/2)(1 - \cos 3\varphi) \quad (8a)$$

$$U(\psi) = (U_\psi/2)(1 + \cos 3\psi) \quad (8b)$$

with the barrier heights of $U_\varphi = 0.6$ Kcal/mol and $U_\psi = 0.2$ Kcal/mol. We also follow Scheraga^{12c)} for the torsional potentials of side chains, which is generally written in a form of

$$U(\chi) = (U_\chi/2) \{1 - (-1)^n \cos n\chi\} \quad (9)$$

The barrier heights U_χ and n for all the side-chain bonds are tabulated in Table II.

Table II. Parameters used for the side-chain torsional potential. The set of (n, U_χ) in eq. (9) is listed along with the name of the rotational bond of each amino acid^{a)}.

| Amino acid | Code | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|------------|------|----------------------------|------------------------------|----------------------------------|------------------------------------|-------------------------------|-------------------------------|
| Cys | C | $C^\alpha-C^\beta(3, 2.8)$ | | | | | |
| Asp | D | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | $C^\gamma-O^{\delta_2}(2, 8.75)$ | | | |
| Glu | E | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 2.8)$ | $C^\gamma-C^\delta(-)$ | $C^\delta-O^{\epsilon_2}(2, 8.75)$ | | |
| Phe | F | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | | | | |
| His | H | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | | | | |
| Ile | I | $C^\alpha-C^\beta(3, 3.1)$ | $C^\beta-C^\gamma_1(3, 2.8)$ | | | | |
| Lys | K | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 2.8)$ | $C^\gamma-C^\delta(3, 2.8)$ | $C^\delta-C^\epsilon(3, 2.8)$ | $C^\epsilon-N^\zeta(-)$ | |
| Leu | L | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 3.1)$ | | | | |
| Met | M | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 2.8)$ | $C^\gamma-S^\delta(3, 1.6)$ | | | |
| Asn | N | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | | | | |
| Gln | Q | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 2.8)$ | $C^\gamma-C^\delta(-)$ | | | |
| Arg | R | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(3, 2.8)$ | $C^\gamma-C^\delta(3, 2.8)$ | $C^\delta-N^\epsilon(-)$ | $N^\epsilon-C^\zeta(2, 20.0)$ | $C^\zeta-N^{\eta_2}(2, 20.0)$ |
| Ser | S | $C^\alpha-C^\beta(3, 0.8)$ | $C^\beta-O^\gamma(3, 1.6)$ | | | | |
| Thr | T | $C^\alpha-C^\beta(3, 3.1)$ | $C^\beta-O^\gamma_1(3, 1.6)$ | | | | |
| Val | V | $C^\alpha-C^\beta(3, 3.1)$ | | | | | |
| Trp | W | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | | | | |
| Tyr | Y | $C^\alpha-C^\beta(3, 2.8)$ | $C^\beta-C^\gamma(-)$ | $C^\zeta-O^\eta(2, 3.5)$ | | | |

a) The free rotation is indicated as $(-)$.

Typical single bonds of aliphatic carbons have three-fold ($n=3$) potentials with relatively high barrier of *ca.* 3 Kcal/mol and have the minima at *trans* ($\chi=180^\circ$) and *gauche* ($\chi=\pm 60^\circ$) rotational positions. No intrinsic potential is allotted to the bonds connecting the aromatic rings (i.e., χ_2 of Phe, Tyr, Trp and His^{9,12a)}). The potentials for the $N^\epsilon-C^\zeta$ and $C^\zeta-N^\eta$ bonds of arginine and the $C^\zeta-O^\eta$ bond of tyrosine are taken from Momany *et al.*⁹⁾ We ignore, for simplicity, the torsional potential related with disulfide bonds, i.e., about the S-S and $C^\beta-S$ bonds of cystine.

The electrostatic energy is only calculated for the dipole-dipole interactions among polar groups, assuming that all the ionizable groups of side chains and of both termini of the protein backbone, such as amino and carboxyl groups, are of non-charged forms to exclude the Coulombic interactions between net charges. The rational of this hypothesis is discussed later. The dipole moment of a polar group is decomposed, according to the monopole approximation, into partial charges

located on the constituent atoms. Then, the dipole interaction is evaluated as the sum of pairwise electrostatic interactions of the partial charges, q ;

$$U_{el} = \sum_{ij} \frac{q_i q_j}{Dr_{ij}} \quad (10)$$

with use of the apparent dielectric constant of $D=4$. Table III shows the partial

Table III. Partial charges of polar atoms in the protein backbone and in the side chains^{a, b)}

| Backbone peptide | | Asn and Gln | |
|----------------------------|--------|--------------|--------|
| N | -0.305 | C γ | +0.449 |
| H | +0.272 | O δ_1 | -0.416 |
| C' | +0.449 | N δ_2 | -0.577 |
| O | -0.416 | H | +0.272 |
| | | H | +0.272 |
| Ser and Thr | | Lys | |
| C β | +0.093 | N ζ | -0.544 |
| O γ | -0.426 | H | +0.272 |
| H | +0.333 | H | +0.272 |
| Tyr | | Arg | |
| C ζ | +0.107 | N ϵ | -0.305 |
| O γ | -0.440 | H | +0.272 |
| H | +0.333 | C ζ | +0.093 |
| Asp and Glu | | N η_1 | -0.305 |
| C γ | +0.529 | H | +0.272 |
| O δ_1 (double bond) | -0.416 | N η_2 | -0.577 |
| O δ_2 | -0.446 | H | +0.272 |
| H | +0.333 | H | +0.272 |

a) Partial charges are given in unit charge, e_0 .

b) The superscript of atoms should be shifted from Table II for Thr, Glu and Gln.

charges¹²⁾ of the backbone and side-chain polar groups. Note that the summation of eq. (10) should cover all the partial charges belonging to the dipole moments considered to fulfil the neutrality condition of the total charges; for instance, some of the atom pairs between the neighboring peptide groups of the backbone, and within the guanidyl group of arginine, always remain at constant distances, but should be included in the computation.

The hydrogen-bond energy has been expressed in various forms in the conformational energy calculations.^{9,12,15)} The hydrogen bond has not only the quantum mechanical nature, but also the strong electrostatic interactions¹⁶⁾ among the proton, H-donor and H-acceptor atoms, interacting within shorter distance than the sum of their van der Waals radii. The situation may be reproduced by the usual electrostatic interactions among partial charges, in combination with a proper repulsive term.¹⁷⁾ As a repulsive term, we simply use the nonbonded energy of eq. (7) between the proton and the H-acceptor atom calculated only for the close disposition shorter than a certain cut-off distance, and assume the zero-potential

beyond it. The cut-off distance of 1.5\AA is chosen for the hydrogen bond of N-H...O, O-H..O and O-H...N types. Fig. 3 shows the resulting potential between the N-H group and the C-O group of two peptides. The potential has a minimum

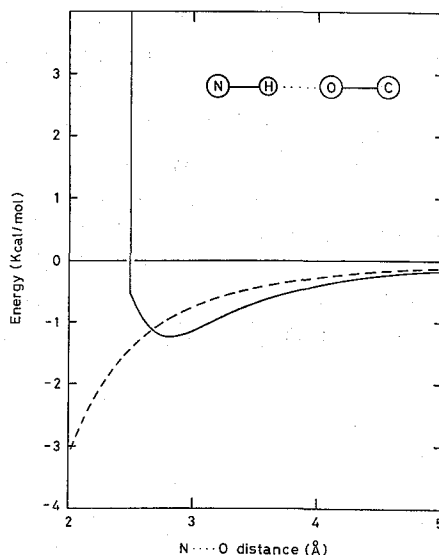


Fig. 3. Hydrogen-bond energy between the NH group and the CO group in the linear disposition, plotted against the N...O distance. The total energy (solid curve) is the sum of the electrostatic interactions (the broken curve) and the nonbonded energies among the four constituent atoms, where the nonbonded potential between the H and O atoms is assumed zero beyond the H...O distance of 1.5\AA .

due to the combination of attractive electrostatic interactions (the dotted curve) and the repulsive force attributed mainly to the nonbonded energy between the N and O atoms. The N...O distance (2.8\AA) and the energy depth (-1.2 Kcal/mol) obtained at the minimum are reasonable (the corresponding values, obtained when the ECEPP energy function⁹) is applied to the same system of Fig. 3, are 3.0\AA and -1.2 Kcal/mol , respectively). The other types of hydrogen bonds yield similar results: the optimum O...O distance and the depth in a system between hydroxyl O-H and the peptide C-O are 2.7\AA and -1.3 Kcal/mol , respectively. The NH groups of imidazole and indole rings may be expected to form relatively weak hydrogen bonds with a polar oxygen. This type of interaction is treated in a manner that the NH group (a united atom) and O atom can approach to each other up to 2.5\AA (=cut-off distance of the nonbonded energy) but with no attractive energy, since partial charges are not assumed on these NH groups (Table III).

When the protein in question has a disulfide bond, the artificial potential of eq. (11) is taken to force the closing of a S-S bond in the correct disposition.^{12c} The pairings of half-cystine residues should be specified in case of several disulfide bonds

existing in a protein.

$$U_{SS} = (1/2)K_{SS}(r_{SS}-r_0)^2 \quad (11)$$

In eq. (11), r_{SS} is a computed S-S distance and r_0 is its standard value, 2.0Å. The force constant K_{SS} must be large enough to make sure the loop closed; a value of 600 Kcal mol⁻¹Å⁻² was taken. Other restrictions, e.g., on the bond angles related with the S-S bond,^{12c)} were not considered to avoid the complexity.

The energy calculation for proline is an exceptional case. We assume some rotational freedom around the N-C^α bond (ϕ) of proline residue, accompanied with the deformation of the pyrrolidine ring, as already stated (Section II). The ring deformation energy, including the bending energy of the bond angles within the ring, the torsional energy around the bonds and the nonbonded energy among the atoms of a proline residue, has been calculated against the change in the dihedral angle ϕ ^{10a)}. The resulting total energy, given in Fig. 4, shows an almost sym-

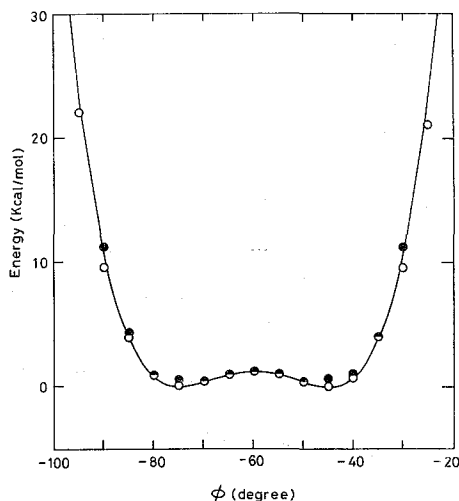


Fig. 4. Ring-deformation potential of proline, represented as a function of the dihedral angle ϕ . The precise evaluation^{10a)} including the bending, torsional, nonbonded energies is indicated by the open and filled circles (corresponding to two types of the deformation.^{10a)}) The solid curve shows the simple approximated function of eq. (12).

metrical pattern centered at $\phi = -60^\circ$. The low-energy region of ϕ ranges from -80° to -40° , and the ϕ angle is in effect forbidden outside of this range.^{10a)} The entire potential can be well fitted to a forth polynomial of ϕ having the minima at $\phi = -45^\circ$ and -75° , as demonstrated in Fig. 4. The function form is

$$U_{\text{pro}}(\phi) = U_{\phi_0}(\phi - \phi_0 + \delta)^2(\phi - \phi_0 - \delta)^2/\delta^4 \quad (12)$$

where ϕ_0 and δ are the constant angles of -60° and 15° , respectively, and $U_{\phi_0} = 1.2$ Kcal/mol. This potential has to be used for the proline ϕ angle instead of

the torsional energy of eq. (8a).

FIRST DERIVATIVE OF CONFORMATIONAL ENERGY

The energetic refinement of a protein structure, starting from a certain initial conformation, is carried out by minimizing the conformational energy to locate the global minimum. We⁵⁾ have used the Davidon's method¹⁸⁾ as a minimizer which utilizes the first derivatives of the energy with respect to all variables of dihedral angles, in our case, to determine the direction of the steepest descent in the conformational space.

The derivatives of the torsional potential and the ring deformation energy of proline, given in Section IV, can be analytically deduced because of their forms expressed in the direct functions of a dihedral angle. The other energies used are, however, all expressed in terms of the distance, r , between an atomic pair. In these cases, a simple method to obtain the first derivative may be a numerical one to calculate the energy difference upon a slight change (Δ) of an independent variable (θ), i.e., $[U(\theta + \Delta) - U(\theta)]/\Delta$, in keeping all of the other variables fixed. This method has been used in our previous studies.⁵⁾ However, we will present, in the following, a more rigorous method, which is also numerical but deals directly with the differential form of an energy function.

Consider an energy function, $U(r)$, expressed with an explicit variable of the distance, r , between two atoms A and B, the relative positions of both of which depend on the rotation, θ , about a bond of P-Q (see Fig. 5). This stands for the

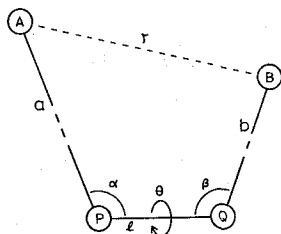


Fig. 5. Schematic representation of the relation between an atom pair of A and B, and the rotation θ about a single bond P-Q. The atoms A and P, and also B and Q, are connected through the polypeptide chain.

general situation for the single atom-pair energy, U , and a dihedral angle, θ , with other variables fixed. Then, the first derivative of U with respect to θ is written as

$$\frac{\partial U}{\partial \theta} = \frac{\partial U(r)}{\partial r} \cdot \frac{\partial r}{\partial \theta} \quad (13)$$

The first term in the multiple of eq. (13) can be derived analytically. In order to get the second term explicitly, the distance r should be expressed in terms of θ . Using the P-Q bond length l , the distance a and b , and the angles α and β defined in Fig. 5, all of which have constant values with respect to the rotation of θ (its zero

rotational position is defined as A and B atoms being *cis* against the P-Q bond), the distance r is written as

$$r = [(a \cos \alpha + b \cos \beta - l)^2 + (a \sin \alpha - b \sin \beta \cos \theta)^2 + (b \sin \beta \sin \theta)^2]^{1/2} \quad (14)$$

Then

$$\frac{\partial r}{\partial \theta} = ab \sin \alpha \sin \beta \sin \theta / r \quad (15)$$

The values of a , b , α , β and θ have to be calculated from the atomic coordinates for each of the atompairs considered.

The general expression of the first derivative of each of the energies with respect to a rotational angle θ_k (anyone of φ 's, ψ 's and χ 's) are given as follows. For the nonbonded energy of eq. (7)

$$\frac{\partial U_{nb}}{\partial \theta_k} = \sum_{ij} 6 \left(\frac{B_{ij}}{r_{ij}^8} - \frac{2A_{ij}}{r_{ij}^{14}} \right) a_{ij} b_{ij} \sin \alpha_{ij} \sin \beta_{ij} \sin \theta_{ijk} \quad (16)$$

where the subscripts of θ_{ijk} denote that this rotation is defined about the same bond of the dihedral angle θ_k but its value depends on the atompair i and j (see above). Similarly, for the electrostatic interaction of eq. (10)

$$\frac{\partial U_{el}}{\partial \theta_k} = - \sum_{ij} \frac{q_i q_j}{D r_{ij}^3} a_{ij} b_{ij} \sin \alpha_{ij} \sin \beta_{ij} \sin \theta_{ijk} \quad (17)$$

and for the loop-closing potential of the S-S bond (eq. 11)

$$\frac{\partial U_{ss}}{\partial \theta_k} = K_{ss}(1 - r_0/r_{ss}) ab \sin \alpha \sin \beta \sin \theta_k \quad (18)$$

It takes about twice as much time to compute these derivatives compared with the simple method already mentioned, because the computation of a , b , α , β and θ , required for each atomic pair, is time consuming. The method presented here is, however, more rigorous than the usual one, so that it may become necessary in case that the precision of the computation is demanded.

DISCUSSION

One way to check how the various energy functions work, all at once, in the application to peptide molecules may be to examine the single-residue energy of amino acids. The energy map of a single residue of L-alanine (with two peptide groups) is shown in Fig. 6, which was obtained by use of the same geometry and the same energy functions as those in this article.¹⁹⁾ The energy map seems reasonable having the global minimum at the extended region in the φ , ψ space, and yielding the low-energy region reached to the right-handed α -helical region through a narrow bridge at around $\psi=0^\circ$. The characteristic ratio of poly-L-alanine estimated from this energy map shows a good accordance with experimental values as well as other theoretical works.¹⁹⁾

Energy Computation of Protein Structure

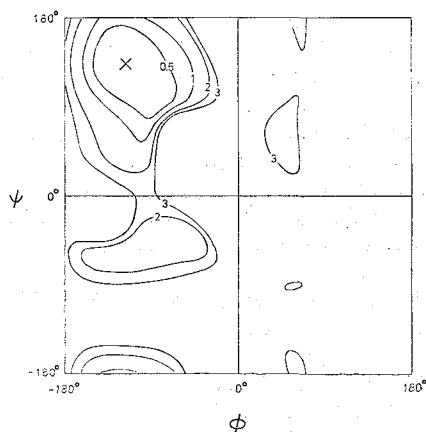


Fig. 6. Energy map of the alanine single residue, N-acetyl-N'-methyl-L-alanine. The energy contours are in Kcal/mol above the global minimum (indicated with an X).

By using the amino acid geometry of Section II and the energy function of Section IV, a method to search the energetically refined structure of a protein, circumventing multiple local minima encountered in the minimization process, was already described in the preceding papers.⁵⁾ The final result for BPTI, including every contribution of the energy terms to the total energy and the root-mean squares deviations from its X-ray structure, will be published elsewhere.^{5c)}

We have made two major simplifications in these studies, as pointed out in Section IV. One is the complete neglect of the solvent (hydrophobic) effects, so that we have assumed as if a protein exists in vacuum (except the use of $D=4$ for the dielectric constant, which reflects the surrounding solvent). A reason of this assumption is that, although the hydrophobic effect is crucial as the stabilization force upon the protein folding,²⁰⁾ it may not be so when the relative stabilities among well-packed conformations are only concerned as in the present case. Another reason lies in the difficulty to evaluate it for a given protein conformation, despite the fact that several attempts have been made intensively.^{7a,21)} Another approximation made in our calculation is to neglect the Coulombic interaction among the net charges. Although this is an unrealistic assumption against a protein in an aqueous solution, the charged groups are generally found to exist at the protein surface²²⁾ so that strong Coulombic interactions between naked charges may be rare to occur within a globular protein. In addition, the interactions between charged groups are weakened by interactions with solvent ions as observed for polyelectrolyte solution.

The inclusion and development of these terms might be necessary in the next step when one proceeds to such a study that concerns the conformational change of a protein.

REFERENCES

- (1) Protein Data Bank (The center in Japan: Institute for Protein Research, Osaka University)
- (2) G.E. Schulz and R.H. Schirmer, "Principles of Protein Structure", Springer-Verlag, New York, 1979.
- (3) (a) M. Levitt and S. Lifson, *J. Mol.*, **46**, 269 (1969), (b) M. Levitt, *J. Mol. Biol.*, **82**, 393 (1974).
- (4) (a) P.K. Warne, N. Go, and H.A. Scheraga, *J. Comput. Phys.*, **9**, 303 (1972),
 (b) P.K. Warne and H.A. Scheraga, *Biochemistry*, **13**, 757 (1974),
 (c) D. Rasse, P.K. Warne, and H.A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **71**, 3736 (1974),
 (d) M.K. Swenson, A.W. Burgess, and H.A. Scheraga, in "Frontiers in Physicochemical Biology", 1977.
- (5) (a) K. Nishikawa and T. Ooi, *J. Phys. Soc. Japan*, **32**, 1338 (1972),
 (b) T. Ooi, K. Nishikawa, M. Oobatake, and H.A. Scheraga, *Biochim. Biophys. Acta*, **536**, 390 (1978),
 (c) T. Ooi and K. Nishikawa, to be published.
- (6) (a) A. Cooper, *Proc. Natl. Acad. Sci. USA*, **73**, 2740 (1976),
 (b) N. Go, *Biopolymers*, **17**, 1373 (1978).
- (7) (a) K.D. Gibson and H.A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **58**, 420 (1967),
 (b) L.G. Dunfield, A.W. Burgess, and H.A. Scheraga, *J. Phys. Chem.*, **82**, 2609 (1978).
- (8) L. Pauling, "The Nature of the Chemical Bond", Cornell Univ. Press, Ithaca, N.Y., 1960.
- (9) F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
- (10) (a) K. Nishikawa and T. Ooi, *Prog. Theor. Phys.*, **46**, 670 (1971),
 (b) W.L. Mattice, K. Nishikawa, and T. Ooi, *Macromolecules*, **6**, 443 (1973).
- (11) C.M. Venkatachalam, B.J. Price, and S. Krimm, *Macromolecules*, **7**, 212 (1974).
- (12) (a) T. Ooi, R.A. Scott, G. Vanderkooi, and H.A. Scheraga, *J. Chem. Phys.*, **46**, 4410 (1967),
 (b) J.F. Yan, G. Vanderkooi, and H.A. Scheraga, *J. Chem. Phys.*, **49**, 2713 (1968),
 (c) H.A. Scheraga, *Adv. Phys. Org. Chem.*, **6**, 103 (1968).
- (13) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).
- (14) M. Oobatake and T. Ooi, *J. Theor. Biol.*, **67**, 567 (1977).
- (15) A.T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
- (16) (a) G.C. Pimentel and A.L. McClellan, "The Hydrogen Bond", W.H. Fieser, San Francisco & London, 1960,
 (b) P.A. Kollman and L.C. Allen, *Chem. Rev.*, **72**, 283 (1972).
- (17) D. Poland and H.A. Scheraga, *Biochemistry*, **6**, 3791 (1967).
- (18) W.C. Davidon, U.S. AEC Research and Development Report, ANL-5990 (1959).
- (19) K. Nishikawa and T. Ooi, *Macromolecules*, **11**, 644 (1978).
- (20) (a) W. Kauzmann, *Adv. Protein Chem.*, **14**, 1 (1959),
 (b) G. Nemethy and H.A. Scheraga, *J. Phys. Chem.*, **66**, 1773 (1962),
 (c) C. Tanford, *Adv. Protein Chem.*, **23**, 121 (1968).
- (21) (a) A.J. Hopfinger, *Macromolecules*, **4**, 731 (1971),
 (b) Z.I. Hodes, G. Nemethy, and H.A. Scheraga, *Biopolymers*, **18**, 1565 (1979).
- (22) (a) C. Chothia, *J. Mol. Biol.*, **105**, 1 (1976),
 (b) P.K. Warne and R.S. Morgan, *J. Mol. Biol.*, **118**, 273 (1978).